

Reply to Barrett: affective neuroscience needs objective criteria for emotions

Ralph Adolphs

Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA

Correspondence should be addressed to Ralph Adolphs E-mail: radolphs@hss.caltech.edu

Abstract

I thank Lisa Barrett for the stimulating and comprehensive article explaining her theory of emotion. In what follows I will summarize what I take to be our points of agreement, my confusions, and suggestions for how to move forward.

Key words: emotion; affect; theory

Where we agree (I think)

I agree that our current criteria for categorizing emotions need to be improved, and that words like ‘anger, fear’ etc. need revision. Constructed Emotion Theory seems to go further than this, however, and at least at times advocates an apparent radical relativism (‘Emotion categories are as real as any other conceptual categories that require a human perceiver for existence, such as “money”’), with which I disagree (see below). However, it may be that this relativism is intended only towards specific emotion categories, not to dimensions of emotion and not to emotion generically, which would bring us closer together.

Surely, we cannot choose to apply the term ‘emotion’ to just anything, otherwise our own articles on emotion might or might not be about emotion (or might be about emotion if read in America, but about something else if read in Europe). I think, then, that Lisa’s article is making a substantive and important point about our need to have better criteria for identifying a state as an emotion—but not proposing that we don’t need criteria at all or can make up any; and with this I agree.

I agree with the article’s discussions of neural reuse, predictive coding and dynamic routing; all these are current theories about brain function for which there is considerable support. I also agree that the brain makes predictions even before birth, since there must be innate priors of some sort (for what it’s worth, I think there are innate emotions, but not innate concepts). So, although one

can quibble about details, I broadly agree with the gist of the article’s discussion about brain function.

I largely agree with the entries in Table 2 that seem to be a key ingredient for the Theory of Constructed Emotion (but I mostly don’t agree with the claims of Table 1, not necessarily because the entries are incorrect, but because they seem far too inconclusive), and I agree that any specific emotion (e.g. fear) does not arise from the activity of a single set of neurons (a view I frankly don’t even understand). Finally, I agree with the facts described about patients with amygdala damage.

The section, ‘Taking a network perspective’, puts forth several interesting hypotheses about the functions of resting-state networks. I agree that brain function depends on distributed networks, and that the current set of networks from resting-state data likely make important contributions to many cognitive functions, but I don’t see this as requiring us to neglect more classical units of analysis such as specific neuroanatomical structures and the systems in which they participate. I think Lisa would also agree with this: so, there are a variety of choices available for how you want to study the brain at the systems level: some might involve resting-state networks, some specific neuroanatomical structures or neuronal populations with them, and none are mutually exclusive.

Where I am confused

I said earlier that I broadly agree with what Lisa’s article says about brain function (predictive coding, neural reuse,

Received: 11 September 2016; Revised: 11 September 2016; Accepted: 11 October 2016

© The Author (2016). Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

degeneracy etc.). On the other hand these features are (i) not at all specific to emotions and (ii) not exclusive (there appear to be instances of predictive coding in the brain, but not everywhere; many functions show degenerative mappings to neuroanatomy, but not all). But I was confused about the relevance of all this, since nowhere in the article was I able to actually find any criteria for what counts as an emotion.

I find many of the interesting claims in the article complex to evaluate because I am unclear either about what the terms mean, or about the reasoning (or both). Here are two quotes that could be useful seed points for future discussion:

'In emotion research, degeneracy means that instances of an emotion (e.g. fear) are created by multiple spatiotemporal patterns in varying populations of neurons. Therefore, it is unlikely that all instances of an emotion category share a set of core features'. I agree with the first sentence, but don't understand how the second follows. This may be because I am unclear on what 'core features' means; the examples given in the article seem obviously wrong. In my view, the 'core features' are functional features (along the lines of the ones listed in Table 2 of my original article).

'What is a brain for? A brain did not evolve for rationality, happiness, or accurate perception. All brains accomplish the same core task ... to efficiently ensure resources... so that an animal can grow, survive and reproduce'.

I think this conflates proximal and distal mechanisms (and I have the same reaction when Joe LeDoux writes of 'survival circuits'). Yes, everything evolved, and some of that evolution is directed towards adaptations for survival and reproduction. But the aggregate long-term adaptive pressure towards survival and reproduction is mediated by many constrained and more proximal mechanisms. So I would disagree that emotions function 'for survival'. They are one piece, amongst many others, that, in part, in a whole organism, in a certain environment, generally aid survival. But their proximal function has much more specific, and more short-sighted goals: specific functions, for example of the sort that evolutionary psychology hypothesizes. We need to understand those specific functions and map them onto specific emotions, not erase their boundaries by saying they all help survival.

A section that was tough going for me was the one titled, 'The computational architecture of the brain is a conceptual system plus pattern generators', because I don't understand Lisa's concept of a concept. For instance, the claim that, 'when the internal model creates an emotion concept, the result is an instance of emotion', would run counter to my view. As I was at pains to point out in my original article, emotion concepts are not emotions. Although the term 'concept' traditionally refers to the mental representations by which we think about something (i.e. semantic knowledge), Lisa's usage appears to be broader. If it means something like 'all the neuronal activity engaged in an emotion state', then of course I would agree that emotions are concepts in this sense—but I am unclear on what such a redefinition of the term 'concept' would actually elucidate. I end up knowing exactly what I did before: as implemented in the human brain, emotion states typically involve predictions and degeneracy and perceptual processes and motor processes and a whole lot more.

The way forward

I appreciate what I think the Theory of Constructed Emotion is trying to achieve. If I understand it right, the big-picture

motivation is twofold: (i) to forge a more holistic, distributed, dynamic view of how emotions are generated in the brain, and (ii) to free us from traditional categories of thinking about emotions (like the terms for basic emotions).

I am very sympathetic with (ii), but I have doubts that (i) is the way to do it. I agree that traditional emotion categories need revision, or even elimination in some cases, and of course I agree that phrenological ways of thinking about brain function as 'in' specific regions isn't how the brain works. But I nonetheless think that we DO need categories of emotions (and/or dimensions), and we DO need to make distinctions between brain systems and their functions. I have the same concern about various versions of enactive/situated cognition: the fact that many of the traditional distinctions draw the wrong boundaries does not mean we should get rid of distinctions and aim towards some conglomerate holism. We should just aim to redraw the boundaries. If anything, I think we need more distinctions and a more fine-grained taxonomy, not fewer. But most importantly, I disagree that one can begin '.....with the structure and function of the brain, and from there deduce what the biological basis of emotions might be'. I think this is squarely impossible, because I think you absolutely need to begin with observation of behavior and derive your categories from there.

I worry about radical versions of relativism that see emotion terms as social constructs like money. In my view, emotions are perfectly objective in the following formal sense: the statement, e.g. 'Ralph is in a state of fear', is true or false independently of whether I believe it to be true or false. I could be wrong about my own emotion state. Science requires objectivity, and if anybody were free to use the word 'fear' as they chose, we could not have a science of emotion.

In terms of how best to derive the emotion categories that a science of emotion requires, I disagree that,

'... it makes no sense to elevate categories for anger, sadness, fear, disgust and happiness to a common ethological framework for comparing humans with other animals, when there is ample evidence from linguistics, anthropology and psychology that these categories do not offer a robust, universal framework

I simply think we should take our definition of these emotion categories from biology and ethology.

There are two domains I omitted in my original article that Lisa's article acknowledges and that indeed need much more discussion. One is interoception, the second is social communication. It is with respect to the latter that I suspect the Theory of Constructed Emotion may be particularly important, since it is indeed true that emotional displays have been co-opted by, and serve a rich and complex role, in social communication—which is highly variable and flexible. It is here that affective neuroscience would transition to social neuroscience and, as with all distinctions, I think this one too is an important one to keep in mind.

The theme of our invited debate was to give our view on how we think neuroscience should study emotion. For me, that is in part a very practical question: what criteria and terms can we agree upon, and what specific framework can we provide for those wanting to do experiments in affective neuroscience. Towards that end, I am grateful for this debate.

Conflict of interest. None declared.